

AMENDMENTS TO THE SPECIFICATION:

Please amend paragraph [0016] as follows:

**[0016]** In accordance with an embodiment of the present invention, a method is provided for identifying clusters in two-dimensional data. The method comprises: (1) generating a two-dimensional histogram characterized by a grid having an x-axis and a y-axis and a selected number of bins in the x-direction and a selected number of bins in the y-direction, and the data comprising  $n$  pairs of points  $(x_i, y_i)$ ,  $i = 1, \dots, n$  such that the histogram comprises fewer bins than points; (2) determining a density estimate based on the bins; and (3) identifying at least one cluster in the data. The method can further comprise generating a smoothed density estimate (e.g., by using a Gaussian kernel estimator algorithm). The method of the present invention [allow] allows boundaries around clusters to be identified (e.g., as polygons) and simplified (e.g., reducing the number of vertices in a polygon while enclosing approximately the same area within the originally-defined boundary).

Please amend paragraph [0043] as follows:

**[0043]** Finding the valleys in a density estimate is a relatively trivial task in [one dimensional] one-dimensional data, but becomes more complicated in data sets of two or more dimensions. In accordance with the present invention, this task is accomplished by linking data points to neighboring points that are "uphill" from them. In other words, for each point in the data set,  $(x_i, y_i)$ , the algorithm 20 looks at the height of the density estimate evaluated at each of the point's neighbors, that is, at all of the other points that are within some radius,  $r$ , of  $(x_i, y_i)$ . Let  $S = \{x_j, y_j\}, j=1, \dots, s$ , be the set of all of the points within a radius,  $r$ , of  $(x_i, y_i)$ . Next, calculate the gradient of the density estimate between the density estimate evaluated at each point in  $S$  and the density estimate evaluated at the point  $(x_i, y_i)$ :

$$g_j = \frac{z_j - z_i}{\sqrt{(x_j - x_i)^2 + (y_j - y_i)^2}}, \quad j = 1, \dots, s,$$

where  $z_k$  is the height of the density estimate evaluated at  $(x_k, y_k)$ . Then,  $(x_i, y_i)$  is linked to the point in  $S$  which has the highest, positive value of  $g$ . In other words,  $(x_i, y_i)$  is

linked to whichever point in its neighborhood is the farthest uphill according to the density estimate. Points that have no neighbors uphill from them are at the tops of local peaks (indicated generally at 36 in Figs. 3 and 7) in the density estimate and they will not be linked to any other points.

Please amend paragraph [0050] as follows:

**[0050]** As can be seen in Fig. 10, the boundaries 38 that were determined in Step 3 follow the outlines of the bins 32 in the data grid 30. This means that boundaries tend to be jagged, like a ladder with many small steps. Also, since the decisions described in Step 3 are made systematically across the grid on a [bin by bin]bin-by-bin basis, the boundaries contain at least one vertex for each bin it traverses. This typically results in boundaries with a large number of vertices; f]. For example, boundaries created on a 64 x 64 grid might contain dozens of vertices.

Please amend paragraph [0061] as follows:

**[0061]** With reference to Fig. 12, the cluster finder algorithm 20 can be implemented in a programmed apparatus 40 comprising a processing device 42 and a memory device 44. The sample of data points from a source 46 such as a flow cytometer can be stored, for example, in the memory device 44. The processing device 42 can be connected to a user interface 48 from which to receive user inputs and to a display 50. The display 50 can generate screens to provide a user with scatter plots, density estimates, as well as cluster and boundary information (e.g., as illustrated in Figs. 2-7) generated in accordance with the algorithm 20. The display 50 can also generate screens to provide the user with menus and other screen options for selectively viewing the scatter plots and clusters. For example, the user can click on a cluster of points in a scatter plot, and the algorithm 20 can process and display the boundary around that particular cluster. The user can request the algorithm 20 to process and display the boundaries of all clusters it finds [simultaneously] simultaneously. The user can change the parameter sigma described above in connection with the density estimate to make the boundaries more tight or loose.